Phonetic contrasts in foreign language perception: a neuropsychological study on Serbian affricates

Kaufmann, Nuria ; Meyer, Martin ; Schmid, Stephan

Abstract: This study addresses the question to which extent phonetic contrasts of a foreign language are perceived more easily by speakers of a native language that shares similar phonetic categories. The focus lies on two postalveolar and two alveolo-palatal affricates of Serbian: [tʃ] (postalveolar, voiceless), [ʨ] (alveolo-palatal, voiceless), [ʤ] (postalveolar, voiced) and [ʥ] (alveolo-palatal, voiced). Swiss-German dialects have the postalveolar voiceless affricate [tʃ] only, while the Rhaeto-Romance variety of Sursilvan has three different affricates, i.e. [tʃ], [ʨ], and [ʥ]. In a EEG experiment using a Multi-Deviant Mismatch Negativity (MMN) paradigm, 15 Swiss-German speaking adults and 15 Rhaeto-Romance speaking adults between the ages of 20 to 30 years were instructed to focus on a random reading while not paying attention to the auditory stimuli. The hypothesis is a significant difference in processing between the two groups: Swiss-German speakers will not be able to reliably distinguish the four Serbian affricates. Rhaeto-Romance speakers on the other hand are expected to be able to distinguish all four affricates as they share three of the four phonetic categories. A significant group-effect was found to corroborate that Rhaeto-Romance speakers process the Serbian affricates differently from the Swiss-German speakers.

Originally published at:
Kaufmann, Nuria; Meyer, Martin; Schmid, Stephan (2010). Phonetic contrasts in foreign language perception: a neuropsychological study on Serbian affricates. In: Schmid, Stephan; Schwarzenbach, Michael; Studer, Dieter. La dimensione temporale del parlato. Torriana: EDK, 425-444.

PHONETIC CONTRASTS IN FOREIGN LANGUAGE PERCEPTION: A NEUROPSYCHOLOGICAL STUDY ON SERBIAN AFFRICATES

Nuria Kaufmann [a], Martin Meyer [a], Stephan Schmid [b]
[a] Institute of Psychology (Division of Neuropsychology), University of Zurich
[b] Phonetics Laboratory, University of Zurich
nuria_kaufmann@access.uzh.ch, m.meyer@psychologie.uzh.ch, schmidst@pholab.uzh.ch

1. ABSTRACT

This study addresses the question to which extent phonetic contrasts of a foreign language are perceived more easily by speakers of a native language that shares similar phonetic categories. The focus lies on two postalveolar and two alveolo-palatal affricates of Serbian: [tʃ] (postalveolar, voiceless), [tɕ] (alveolo-palatal, voiceless), [dʒ] (postalveolar, voiced) and [dʑ] (alveolo-palatal, voiced). Swiss-German dialects have the postalveolar voiceless affricate [tʃ] only, while the Rhaeto-Romance variety of *Sursilvan* has three different affricates, i.e. [tʃ], [tɕ], and [dʑ].

In a EEG experiment using a Multi-Deviant Mismatch Negativity (MMN) paradigm, 15 Swiss-German speaking adults and 15 Rhaeto-Romance speaking adults between the ages of 20 to 30 years were instructed to focus on a random reading while not paying attention to the auditory stimuli. The hypothesis is a significant difference in processing between the two groups: Swiss-German speakers will not be able to reliably distinguish the four Serbian affricates. Rhaeto-Romance speakers on the other hand are expected to be able to distinguish all four affricates as they share three of the four phonetic categories.

A significant group-effect was found to corroborate that Rhaeto-Romance speakers process the Serbian affricates differently from the Swiss-German speakers.

2. INTRODUCTION

There is a diversified discussion on how and when we best learn a foreign language (L2). Some advocate that foreign-language learning is no longer possible without any accent after a ‘Critical Period’ (e.g., Lenneberg, 1967; Kuhl, 2004). Others plead in favor of a continuous mode of foreign-language learning which does not differ significantly between children and adults (e.g. Friederici, 2005). The Critical Period Hypothesis states that an L2 exhibits different processing patterns than the L1. A ‘less is more’ Hypothesis on the other hand states that processing patterns could be similar, provided the new grammar to be learnt is small (Friederici *et al.*, 2002). This would conform to the assumption that language competence in the L2 affects processing patterns more significantly than age of acquisition (e.g. Winkler *et al.*, 1999).

A number of neuropsychological studies reveal an improved ability to discriminate foreign language sounds with higher language proficiency (e.g. Winkler *et al.*, 1999). Mismatch negativity paradigms have shown that fluent non-native speakers develop a cortical memory for the foreign language phonemes (Näätänen *et al.*, 1997 and Winkler *et al.*, 1999). Such recognition patterns presumably develop gradually with the exposure to the new language. Even in a well-learned second language, however, phoneme representations of the native language were found to exert a strong influence on contrast detection

(Nenonen *et al.*, 2005). Consequently, different mother tongues (L1s) could out-fit one differently to learn a certain foreign language. Thus, we consider the MMN approach most suitable to address our question at issue.

The mismatch negativity (MMN) is a negative deflection that peaks approximately 100-250 ms after the stimulus onset. Classically, the MMN is elicited in the so-called 'oddball paradigm' (see below) as response to sudden changes (deviants) in an auditory sequence (usually represented by standard stimuli). The MMN is understood as a pre-attentive, automatic response. Nevertheless, its amplitude can be enhanced under attention (Näätänen *et al.*, 2004).

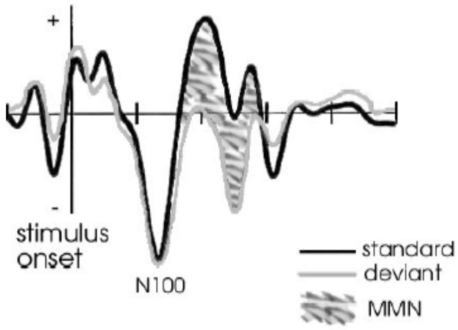

Figure 1: Schematic illustration of the oddball paradigm and the resulting MMN (Lipski, 2006: 45)

The MMN can be observed in the difference wave that is obtained by subtracting the Event-related Potential (ERP) of the standard-stimulus from the deviant-ERP. As the MMN arises only as a response to a new event, the formation of a memory trace to the standard stimuli is preconditioned (Titova & Näätänen, 2001). Incoming deviant stimuli are compared to the regular pattern of the standard sound. Näätänen and colleagues (1997) find enhanced MMN responses to phoneme changes that are relevant to the subject's native language. Reflecting the processing of abstract regularities, long-term memory traces and learning effects, MMN is therefore applicable for the study of cognitive functions.

Up to now, MMN experiments have been conducted on various syllable types (compare Näätänen *et al.*, 1997; Lipski, 2006), but the considerable variety of affricate categories across the languages of the world (Ladefoged & Maddieson, 1996: 90-91) calls for advanced research also on this specific topic. Let us therefore briefly illustrate the affricate subsystems of the three languages involved in the present study, i.e. Serbian, Rhaeto-Romance (Sursilvan), and Swiss-German.

The consonant inventory of Serbian is rather complex (Corbett, 1987: 396). Within the manner of articulation of affricates, Serbian differentiates four categories that are used in our experiment, namely [ʧ] (postalveolar, voiceless), [ʨ] (alveolo-palatal, voiceless), [ʤ] (postalveolar, voiced), and [ʥ] (alveolo-palatal, voiced); it has also has [ts] (alveolar, voiceless) which is not part of the experiment. Phoneticians disagree, however, whether the palatal obstruents are affricates or stops; moreover, they are sometimes described as palatal, sometimes as alveolo-palatal (Morén, 2006).

The Rhaeto-romance language territory in Switzerland is divided into five dialects (Haiman & Benincà, 1992): *Sursilvan, Sutsilvan, Sumiran, Puter* and *Vallader*. For our experiment, *Sursilvan* has been chosen, because it is the most spoken dialect. *Sursilvan* shares three of the four affricates with Serbian, namely [tʃ], [tɕ] and [dʑ] (Liver, 1999). Again, scholars disagree with regard to the phonetic description of [tɕ] and [dʑ], which are classified either as stops or as affricates on the one hand, and as palatal or as palato-alveolar on the other (compare Brunner, 1963; Schmid, 2010). For the purpose of this study, we consider them as alveolo-palatal affricates, just as the Serbian ones. *Sursilvan* [tʃ] also seems to share the typical lip rounding of Serbian (Morén, 2006), but it lacks the voiced postalveolar affricate [dʒ] that exists in Serbian (as well as in Italian and in the Rhaeto-Romance dialects of the Engadine).

Now turning to Swiss-German, we might illustrate its consonant inventory by referring to the Zurich dialect, which contains four voiceless affricates, namely labial [pf], alveolar [ts], postalveolar [tʃ], and velar (sometimes uvular) [kx] (Fleischer & Schmid, 2006). Thus, out of the manner of articulation we are interested in, Swiss-German only has [tʃ]. Considering that the four affricates in our study differ in voicing and place of articulation, it must be pointed out that Swiss-German speakers do not differentiate contrasts of the affricate category in either dimension. However, due to some knowledge of English (which has both a voiceless and a voiced postalveolar affricate), the distinction of voicing might be easier for them than the detection of another place of articulation.

Affricates	post-alveolar (voiceless)	alveolo-palatal (voiceless)	alveolo-palatal (voiced)	postalveolar (voiced)
Serbian	[tʃ]	[tɕ]	[dʑ]	[dʒ]
Rhaeto-Romance	[tʃ]	[tɕ]	[dʑ]	–
Swiss-German	[tʃ]	–	–	–

Table 1: Affricates in Serbian, Rhaeto-Romance and Swiss-German
(comparison of the phoneme categories that are relevant in this study)

3. MATERIALS AND METHODS

3.1 Stimuli

The four Serbian syllables [tɕa], [dʑa], [dʒa], and [tʃa] served as stimuli in the Electro-Encephalogram (EEG) recording (see 3.3). The usage of CV (consonant-vowel) syllables was motivated by the fact that isolated affricates, especially voiceless ones resemble nonspeech noise. This impression is reinforced by the repetitive presentations that are necessary in EEG experiments. Furthermore, the transitions to subsequent vowels may provide important perceptual cues for the identification of the affricate. Because the vowel [a] is universally unmarked, we decided to apply this vowel. In contrast to [u] and [o], [a] does not lead to anticipatory lip rounding during the production of the affricate and there is no coarticulatory influence of a palatal glide for [a].

The stimuli used for the experiment were digitally recorded in a sound proof chamber at the Phonetics Laboratory of the University of Zurich. A sampling rate of 44100 Hz and 16 bit quantization were used. A female native speaker of Serbian read the four syllables aloud in twelve variations each: they were spoken three times in a CV sequence, in a VCV sequence and in an existing Serbian word (*ćaskati* "to chat", *čarapa* "sock", *đavol* "devil", *džaba* "frog"). All of the syllables that served as acoustic stimuli were pronounced inside a carrier phrase where the preceding segment was a vowel (*Prvo ća, drugo ća, treće ća* "first ća, second ća, third ća"), which allowed us to precisely detect the starting point of the consonant under examination.

3.1.1 Acoustic analysis

In a first step, the duration of the closure and the release phase of 48 affricates was measured manually on the basis of an introspection of the wave forms and spectrograms provided by *Praat* (Boersma & Weenink, 2009); after that, the duration of the whole syllable was noted. In order to guarantee a certain reliability of the measurements, the procedure was repeated in order to obtain two times 24 tokens, including four stimuli – each in three repetitions (see carrier phrase) and two conditions (CV and VCV).

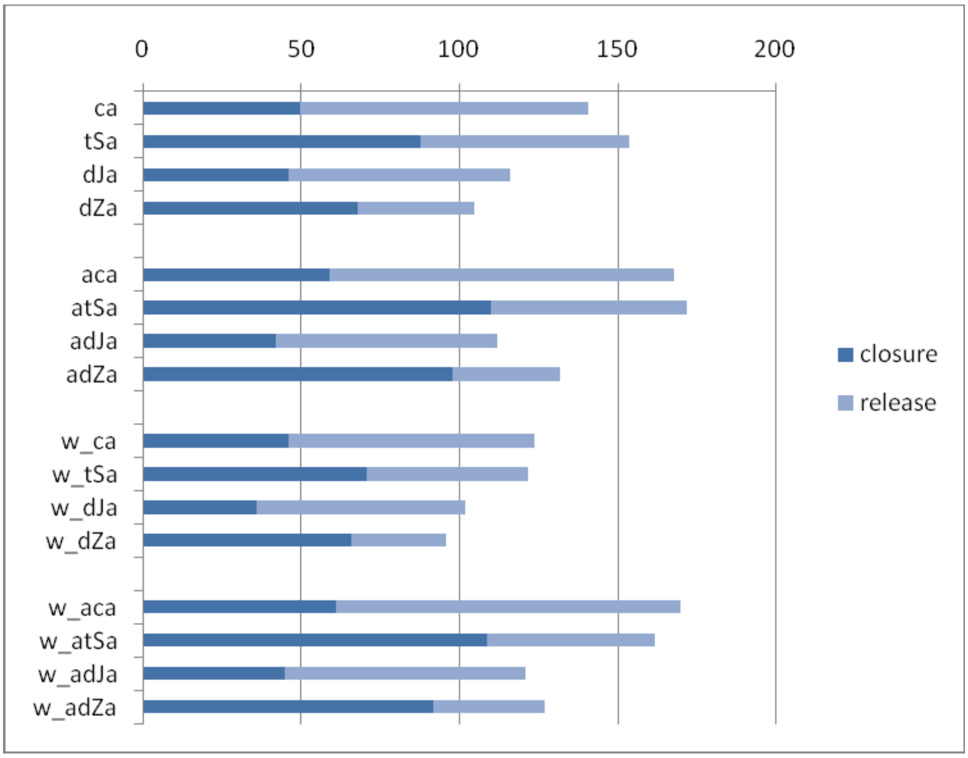

Figure 2: Mean duration values (ms) for the closure and the release phase of the affricates in logatomes (up) and Serbian words (down)

Figure 2 illustrates the mean duration values (ms) for the closure and the release phase of of the affricates in the recorded career sentences. The upper part of the graph shows the mean duration values of the affricates [ʨ], [tʃ], [ʤ] and [ʥ] pronounced in CV and CVC logatoms;[1] the lower part shows the mean duration of the same affricates pronounced in Serbian words.

In all four contexts, voicing clearly affects duration, since the two voiceless affricates are always longer than the voiced ones. As regards place of articulation, it results that the two alveolo-palatal affricates always display a relatively shorter closure phase and a longer release phase than the two postalveolar affricates.

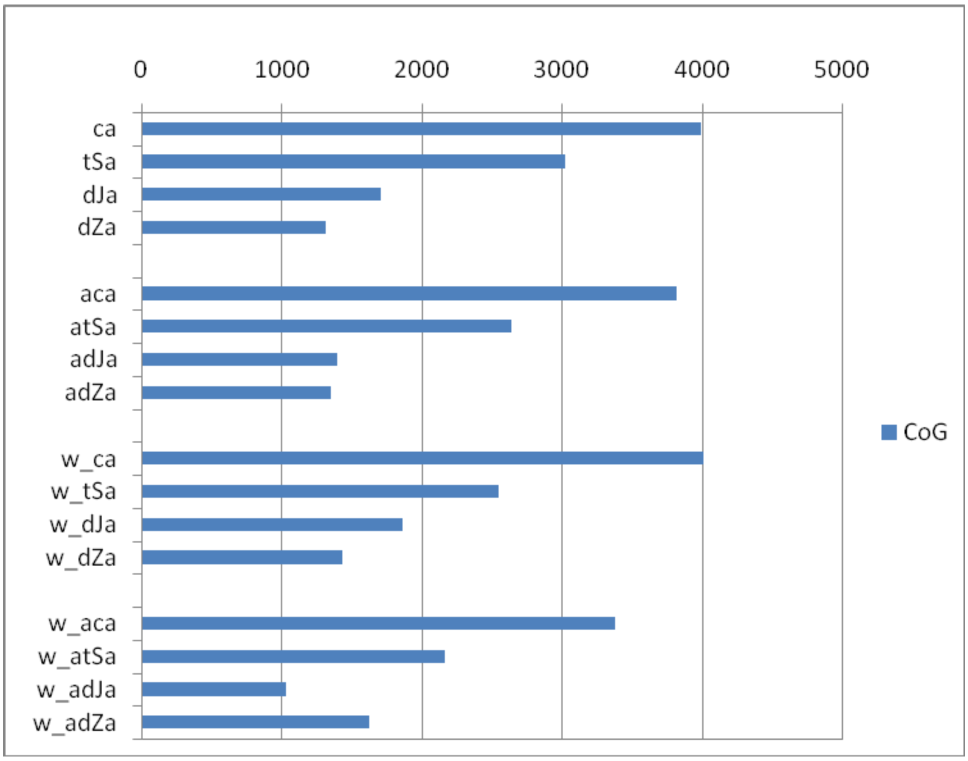

Figure 3: Mean values of the Centre of Gravity (CoG) of each affricate (Hz)

In order to obtain a measure for the spectral characteristics of the affricates, the 'Centre of Gravity' (CoG) was calculated (Forrest *et al.*, 1988; Gordon *et al.*, 2002), using the apposite function in *Praat* (compare Mele & Schmid, 2009: 365-367). Figure 3 illustrates the mean CoG values for the affricates [ʨ], [tʃ], [ʤ] and [ʥ] according to the four different phonosyntactic contexts. The results indicate a significant effect of voicing on the Centre of

[1] For technical reasons, here [ʨ] and [ʥ] are referred to by the symbols [c] and [dJ]; similarly, [tS] stands for [tʃ] and [dZ] for [ʤ].

Gravity, given the clearly higher values for the voiceless affricates; obviously, this result reflects the additional presence of energy in the lower frequency range which appears in the spectrum of voiced obstruents. As regards place of articulation, we note a higher CoG for the palatal affricates as opposed to the postalveolar ones. For the time being, we limit ourselves to observe this as an acoustic fact (with a possible auditive effect in the acute-grave dimension), without speculating on the articulatory nature of the sounds involved; possibly, lip rounding is at stake here.

3.1.2 Selection and editing

The four stimuli used in the experiment were selected from the second recording according to the following criteria: Duration for affricate and vowel about 150 ms, even, constant fundamental frequency (F0) trend. Editing included stylizing the pitch using *Praat* 5045 (Boersma & Weenink, 2009) and setting the overall intensity to 70 dB. Normalization was done using the software *Audition.*[2] This did not change the intensity relation between affricates and vowels in the individual syllables. A Butterworth filter was applied as low-pass filter (5000 Hz) to cut background- and click-sounds using *Audition.* At the onset and at the end of the syllables a smooth rising/falling ramp with duration of 10 ms was added (Gaussian filter). F0 was set to a constant value throughout the vowel with respect to initial F0 value. Duration was normalized by clipping the affricate onset and vowel offset so that each syllable had duration of between 120-185 ms. Finally, the vowel of the syllable [tʃa] was stabilized at a length of 92 ms and was used for all four syllables.

The last step was done in full awareness of the loss of information that is provided by the specific transition of the affricate to the following vowel. As described in Recasens & Espinosa (2007: 149), "the duration of the vowel preceding the affricate ought to be strongly related to the duration of the entire affricate and of its closure period. Spanish, English and Italian data reveal indeed that vowel duration compensates for affricate and closure duration but less clearly so or not at all for frication duration, i.e. the vowel shortens as the affricate and its closure period lengthen and vice versa".

Indeed we observed the same effect. In addition to the described dependency of stimulus-length, the formant constellation in the transition from the affricate to the vowel varies between the four affricates. We performed a behavioural pre-experiment with Swiss-German speakers which showed that the isolated syllables are much too easy to distinguish with this information included. Subjects reported they would be able to easily differentiate the stimuli paying attention only to the 'higher' and 'lower' sounding vowels. However, we were interested in their ability to perceive the spectral part of the affricate only. This confirmed the necessity of taking away this stimulus-specific attributes, although it retrenches the naturalness of our stimuli.

After the final editing, three Serbian and three Rhaeto-Romance speakers were asked to judge the syllables for their 'naturalness' and their discriminability (e.g. Nenonen *et al.*, 2005). Serbian speakers could reliably ascribe each syllable; Rhaeto-Romance speakers encountered increased difficulties, yet they clearly made out "three or more" different syllables.

[2] http://www.adobe.com/products/audition/

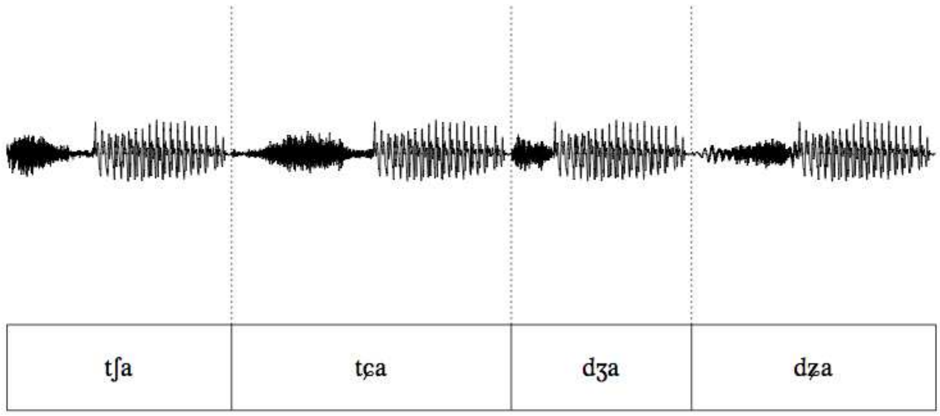

Figure 4: The four Serbian affricates used for the CV stimuli

3.2 Subjects

For the experiment, 30 subjects were recruited: 15 Surselvan mother tongue speakers for the Rhaeto-Romance group and 15 Swiss-German natives. Only righthanded subjects between 20-30 years were assessed. During the installation of the electrodes participants were asked to fill in a questionnaire to file their details: first and foremost their language background (bilingual, second language abilities, etc.). All subjects of both groups learnt English and French in school. Some had knowledge of Spanish. Only four subjects knew Italian. Other second languages were Norwegian (1), Swedish (1), Arabic (2), Hebrew (1) and Latin (3). Knowledge of various second languages might also promote orthographic knowledge which is assumed to influence speech perception (Lipski, 2006). Their contact details were noted in case of further questions.

3.3 Procedure

During the EEG experiment, subjects were seated in an electrically shielded and acoustically attenuated chamber. The data were recorded using a Biosemi active-two amplifier system. 64 active electrodes were installed according to the 10/20 electrode system (Jaspers, 1958) (see figure 5 below).[3] The sampling rate was 512 Hz and impedance was kept below 40 k∧;[4] vertical and horizontal eye movements were recorded by two bipolar channel pairs placed above and below the left eye, and on the outer canthi of both eyes. For off-line re-referencing, an electrode was attached to the tip of the nose. For head and body movements, participants were monitored through a close-circuit camera system. The whole experiment, including welcoming and hair washing lasted two hours. Subjects received 20 Swiss francs for their participation.

[3] http://www.biosemi.com/headcap.htm
[4] http://www.biosemi.com/faq/shielding%20vs%20active%20electrodes.htm

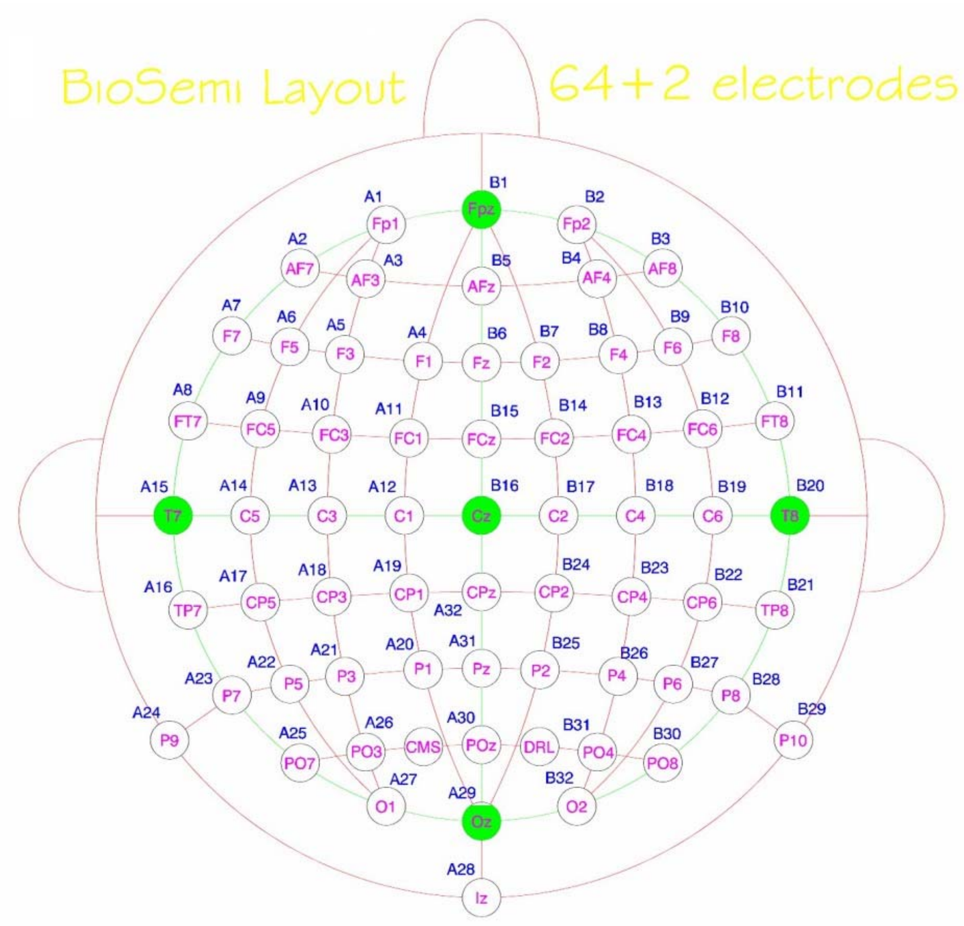

Figure 5: 64 channels 10/20 – layout

3.4 Oddball paradigm

The paradigm follows the idea of Näätänen *et al.* (2004)'s Optimal 1 paradigm. Three 'deviants' (randomly alternating stimuli) are presented alongside the 'standard' (a stimulus which is repeated continuously every second) and not compared individually against the standard as in the classic oddball paradigm. We used an oddball paradigm with 50 percent standard (e.g. [tʃa]) and 50 percent deviant ([ʨa], [ʥa] and [ʤa]) proportion. Furthermore, we used a Multiple-Deviant Paradigm which means that every deviant once acts as the standard. The Inter-stimulus Interval (ISI) was set to 750 ms and the Stimulus Onset Asynchrony (SOA) of 400 ms was jittered. The first two minutes were recorded for closed and open eye-movements (resting EEG). Thereafter, eight passive listening blocks followed. Block sequences were randomized between subjects. Participants were asked to read a random text and not pay attention to the syllables they heard through the headphones, but to treat them as 'background music'. At the beginning of each block there were 15 repetitions of the standard to attune the subjects' ear to the respective standard. Therefore, a stimulus block included 150 deviants and 165 standards; in total, 1200 deviant repetitions and 1320 standard repetitions were used, whereof each of the four stimuli appeared 330 times as a standard and 300 times as a deviant. Between blocks, subjects could recess for as long as they wished. On average, breaks lasted three minutes.

3.5 Data Analysis

The data were analyzed using *BrainVision Analyzer 1.05.0000* and *eegLab 6.01* (Matlab). EEGs were offline treated with a 24 dB zero-phase bandpass-filter from 0.1 to 30 Hz. Channels that displayed changes exceeding 150 ⌠V were discarded for further analysis. Unfortunately, we could not use the nose as reference, as the coordinates of this electrode are unknown to the *eegLab system.* Common average Reference (CAR) was therefore applied.

Eye blinks and horizontal movements were corrected by means of independent component analysis (ICA). Due to technical problems while recording, six subjects (three Rhaeto-Romance and three Swiss-German) had to be discarded. EEG recordings were segmented into 600 ms epochs (100 ms pre- and 500 ms post-stimulus) and averaged for each stimulus type separately with 100 ms pre-stimulus as a baseline. ERPs for all stimuli (each stimulus type as a standard and as a deviant) were averaged for each subject and grand-averaged across subjects.

MMN difference waves were computed by subtracting ERPs to the standard from ERPs to the deviant of a chosen stimulus and grand-averaged. Being able to directly compare the response to a certain stimulus acting both as a standard and as a deviant is one of the main advantages of the Multiple-Deviant Paradigm (compare also Grimm *et al.*, 2008). Peak-detection was carried out over a time-window of 180 ms (120-300 ms after stimulus onset).

The presence of the MMN was statistically verified using analysis of variance, one-sampled and independent *t*-tests with *SPSS* at a significance level of 0.05. Analysis involved comparison of groups, stimuli, peaks and latencies. To verify the existence of a true MMN component, activations at Fz were compared with supra-temporal electrodes (TP9 and TP10).[5]

4. RESULTS

Deviant-related MMN potentials were measured by subtracting ERPs elicited by the stimulus operating as a standard sound from ERPs elicited by the same stimulus operating as a deviant sound. This allowed a direct comparison of the physically identical stimulus differing only in its probability of occurrence. ERPs showed orderly N1 and P2 components at central Cz electrode, comparing the two groups (see below Figure 6) and comparing the four stimuli acting as standards for both groups (compare Figure 7 below). Normal distribution was assured with a Kolmogorov Smirnov test.

[5] For position of electrodes see Figure 5.

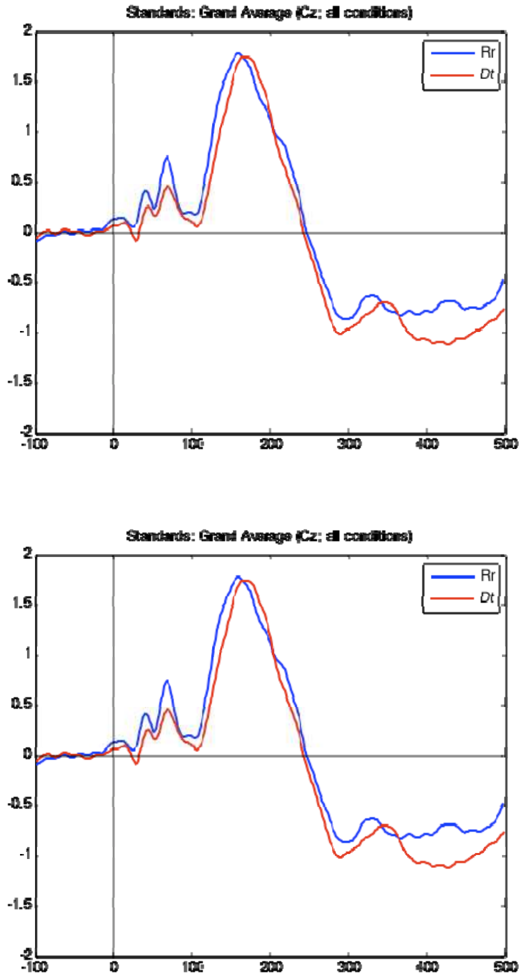

Figure 6: Central Acoustic Event-related Potential (CAERP) plotted at Cz electrode for Rhaeto-Romance speakers (Rr - blue) vs. Swiss-German speakers (Dt - red)

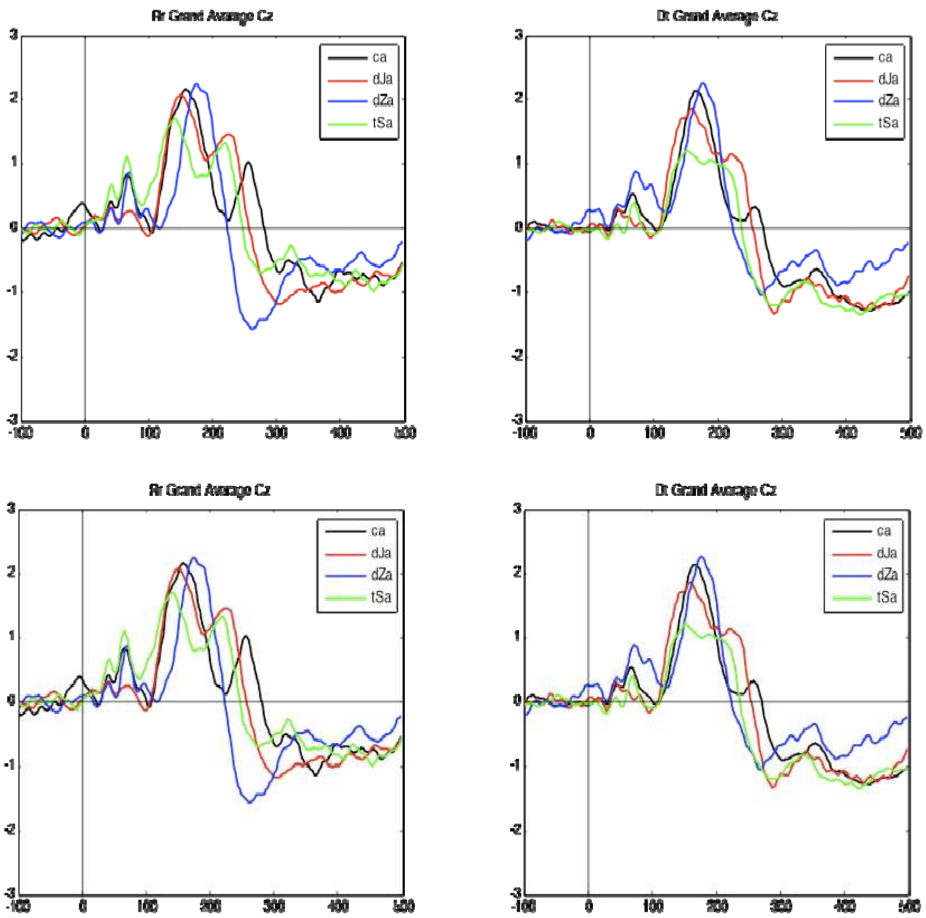

Figure 7: CAERP plotted at Cz electrode
for Rhaeto-Romance speakers (left) vs. Swiss-German speakers (right)[6]

A repeated-measures ANOVA was performed for peaks and latencies separately. The ANOVA included the between-subject factor ‘Group’ (Rhaeto-Romance vs. Swiss-German), ‘Stimulus’ ([tʃa], [tɕa], [dʑa] and [dʒa]) and the within-subject factors ‘Peak’ or ‘Latency’ (three peak or latency values per stimulus, representing the three deviant conditions). For both language groups in all deviant conditions, negative peaks were observed in the deviant-minus-standard difference waves. The comparison ‘Group’ x ‘Stimulus’ x ‘Peak’ revealed a main effect ‘Group’ (p = 0.03) (compare Table 2 below). As expected, the comparison ‘Group’ x ‘Stimulus’ x ‘Latency’ revealed no main effect.

[6] In the graph, [tɕ] and [dʑ] are referred to by the symbols [c] and [dJ]; similarly, [dZ] stands for [dʒ] and [tS] for [tʃ] ([tɕ] = black, [dʑ] = red, [dʒ] = blue and [tʃ] = green).

		Tests of Between-Subjects Effects (ANOVA)	
		df	Sig. (2-tailed)
Interaction Group:		1	0.027*
Stimulus*Peak*Group:	(Greenhouse-Geisser)	5	0.010***

Table 2: ANOVA: comparing 'Group', 'Stimulus' and 'peaks'

Figure 8 provides the mean scores of each group for all stimuli in the respective deviant conditions; standard deviations are marked with bars and amplitude peaks are compared for each stimulus acting in different deviant positions as compared to acting as a standard. Significant differences that showed in the independent samples t-test are marked with asterisks ($p < 0.001$ = ***, $p < 0.01$ =**, $p < 0.05$ = *); no significant differences were found for the comparison of latency means (see figure 9 below).

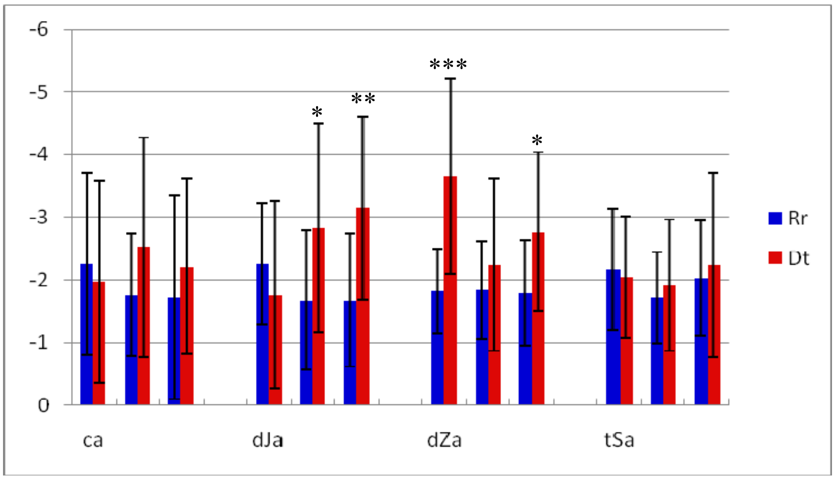

Figure 8: Comparison MMN peak values (µV) and standard deviations
in respective deviant conditions –
Rhaeto-Romance speakers (Rr: blue) and Swiss-German speakers (Dt: red)[7]
(*** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$)

[7] In the graph, [tɕ] and [dʑ] are referred to by the symbols [c] and [dJ]; similarly, [dZ] stands for [ʤ] and [tS] for [tʃ]. The ascribed stimulus represents the standard that was compared to the same stimulus when acting as a deviant.

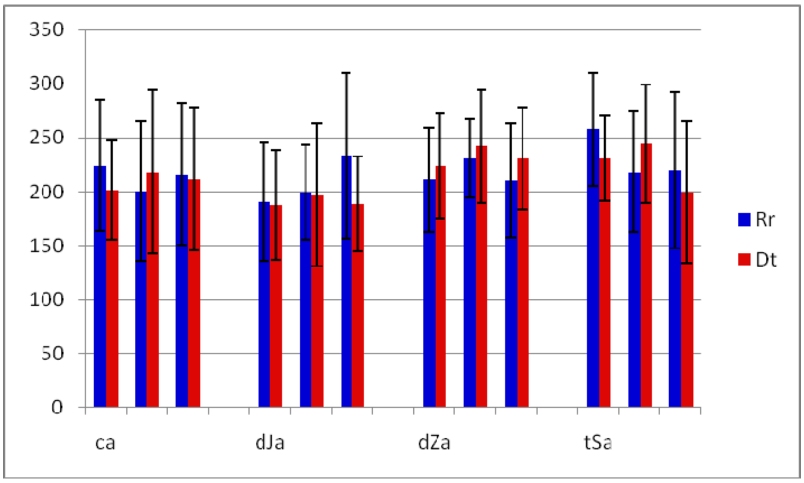

Figure 9: Comparison of groups of MMN latency values (ms) and standard deviations in respective deviant conditions

As expected, the comparison Group * Stimulus * Latency revealed no main effect. Due to slightly different stimulus length (up to 65 ms difference), a systematic latency effect was anticipated. As expected, Tests of Within-Subjects Effects indicated a significant effect for Stimulus (Greenhouse-Geisser $p = 0.013$; not shown in the table), but no interaction with Group.

One-sampled *t*-Tests in both groups for the comparisons of both peaks and latencies were all significant on the $p < 0.001$ level (not shown in the table). Independent Samples *t*-Tests show, that Rhaeto-Romance speakers process phonetic contrasts significantly differently. Surprisingly, the stimulus [tɕa] elicited no significant group difference. Stimulus [ʥa] was processed significantly differently if it served as a deviant beside the standard [tʃa] ($p = 0.01$) and [ʤa] ($p = 0.03$) compared to acting as a standard. Stimulus [ʤa] also displayed significant differences between Rhaeto-Romance and Swiss-German speakers when serving as a deviant in standard blocks [ca] and [tSa] (compare Figure 7 above). Interestingly, these four significant MMN amplitudes are higher for Swiss-German speakers than for Rhaeto-Romance speakers. Based on previous studies with native and non-native phonological contrasts (e.g. Winkler *et al.*, 1999; Peltola *et al.*, 2003; Näätänen *et al.*, 1997), an enhanced MMN for the native-like stimuli was anticipated for the Rhaeto-Romance group. However, in nine out of the twelve contrasts, Swiss-German speakers attained bigger amplitudes than Rhaeto-Romance speakers, four of which were significant (compare Figure 7 above). Surprisingly, no significant group differences were found for the stimulus [tɕa]. As expected, no differences in processing were found for the stimulus [tʃa]. This stimulus is common to speakers of both language groups and should therefore not provoke a significant difference in processing.

The Figures 10 and 11 show the Difference waves Deviant-Standard at Fz electrode for Rhaeto-romance subjects and Swiss-German subjects, respectively. Each stimulus is presented as a standard (blue line) alongside the three different deviants (red lines – left to right) and the resulting difference waves (black lines). The standards from the 1st to the 4th row: [tɕa], [ʥa], [ʤa] and [tʃa].

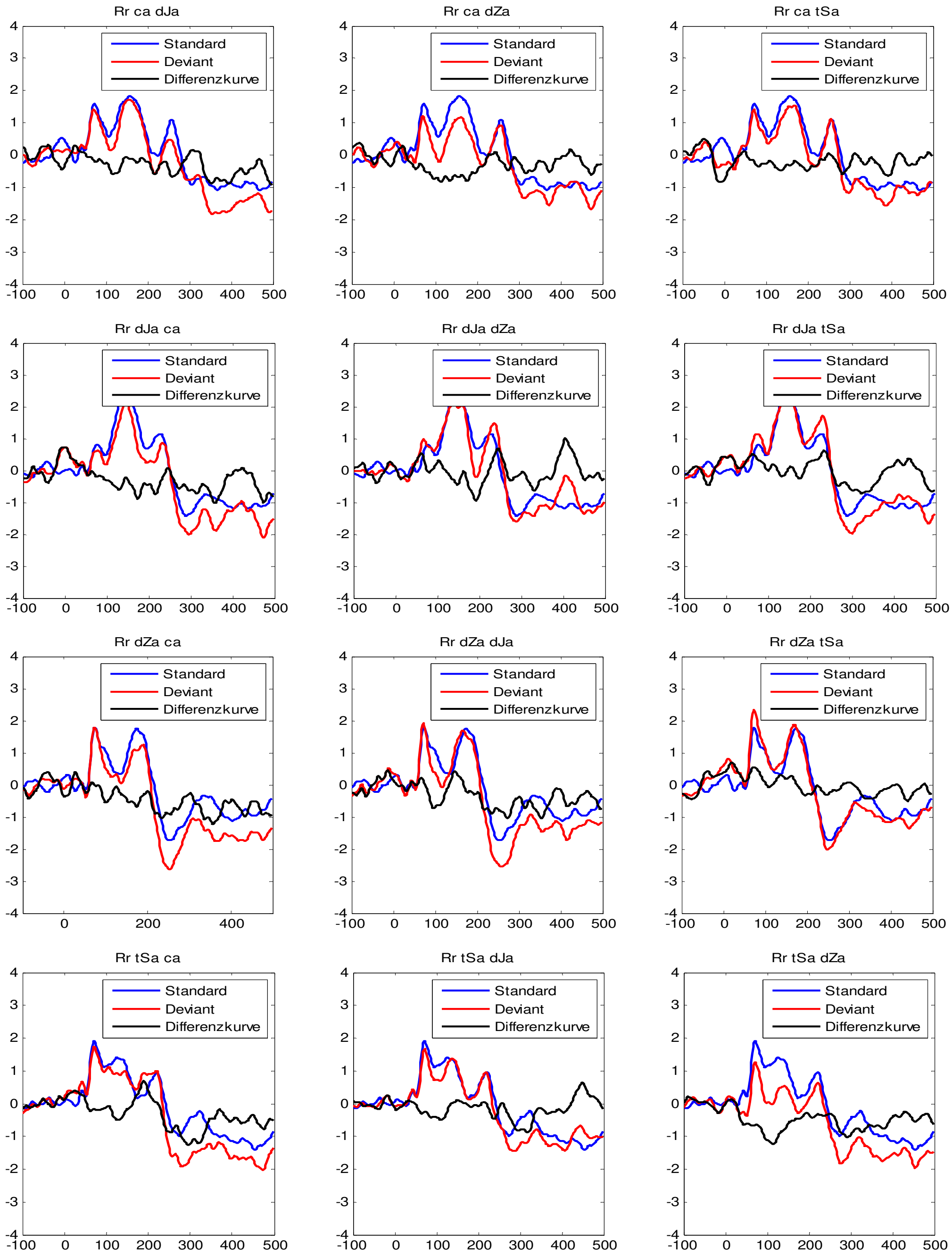

Figure 10: Difference waves Deviant-Standard at Fz electrode for Rhaeto-romance subjects

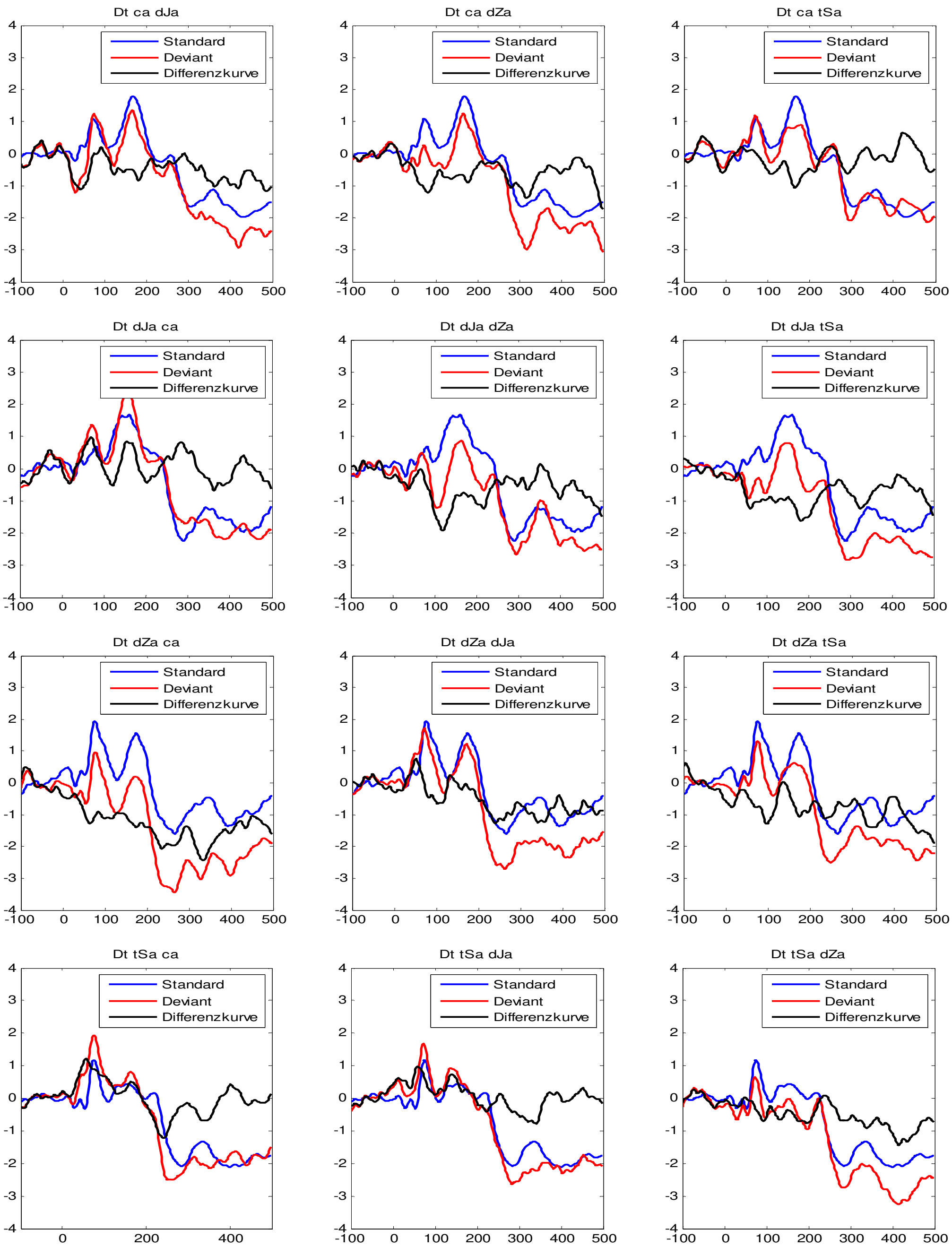

Figure 11: Difference waves Deviant-Standard at Fz electrode for Swiss-German subjects

All MMN curves displayed a typical fronto-central maximum (Fz) with a polarity inversion at the mastoid leads (TP8 & TP7) and latencies between 180-260 ms. The four stimuli are compared in three deviant conditions. For the difference waves, each stimulus is compared in its function as standard to its respective function as a deviant. For example, in the first row of figure 10 (related to the Rhaeto-Romance subjects) and figure 11 (related to the Swiss-German subjects), difference waves of the standard [ʨa] are computed for [ʨa] functioning as a deviant in the first, second and third condition (left to right).

5. DISCUSSION AND CONCLUSION

The overall goals of the MMN experiment were two-fold: first and foremost, to examine the implications of the different language-backgrounds of the two groups, and second to test whether place of articulation or voicing had a stronger influence on the perception of a foreign language phonetic contrast. A significant difference ($p = 0.03$) in MMN amplitudes between groups confirmed a varying way of processing. Half of the expected phonetic contrasts yielded a significant difference between groups. The direction of the effect, however, is unexpected.

With respect to the second goal, we observe a number of interesting findings. What is more, these results challenge the traditional interpretation of the results to the first goal. Instead of the Rhaeto-Romance group, the Swiss-German group shows higher amplitudes. Higher amplitudes in discriminating phonetic contrasts were previously associated with native-like or more proficient processing (compare e.g. Winkler *et al.*, 1999; Näätänen *et al.*,1997). As Rhaeto-Romance shares similar phonetic categories with Serbian, the speakers of this group were expected to out-perform the Swiss-German speakers. We expected higher discrimination ability that would yield larger amplitudes for the Rhaeto-Romance group.

It was hypothesized, that Rhaeto-Romance speakers would be able to differentiate all three stimuli [ʨa], [ʥa] and [ʤa] unknown to Swiss-German speakers. Amplitude differences were expected for contrasts that involved known deviants compared with contrasts that employed an unknown syllable: the Rhaeto-Romance group was expected to show high discrimination amplitudes for deviants [ʨa] and [ʥa] and a lower amplitude for deviant [ʤa]. Except for the deviant [tʃa], lower discrimination amplitudes were expected for the Swiss-German group. Surprisingly, Swiss-German speakers showed higher amplitudes on four out of twelve contrasts that reached significance, even though the contrasts involved deviants that were unknown to them.

Stimulus [ʥa] is known to Rhaeto-Romance speakers. Phoneticians might disagree on the correct phonetic categorization of the Rhaeto-Romance equivalence of the phoneme; nevertheless, the palatal place of articulation as well as high degree of voicing are represented in the phoneme category of interest (not so in Swiss-German, where neither feature is present).

Stimulus [ʤa], on the other hand, is unknown to both Rhaeto-Romance and Swiss-German speakers. Again, both features are known to Rhaeto-Romance speakers. Swiss-German speakers are familiar with the postalveolar affricates, but not with the voiced ones. Possibly, Swiss-German speakers could distinguish the voiced affricates due to previous encounters with English, French and/or Italian.

Stimulus [tʃa] was not expected to reach significance in processing between the two groups as it forms part of the Rhaeto-Romance and the Swiss-German phonetic system alike.

Stimulus [tɕa] might not have yielded a significant result for either group because of too much stimulus editing. It might have been equally difficult for Rhaeto-Romance speakers even though Rhaeto-Romance speakers were able to distinguish the contrast behaviourally which could be interpreted as enhanced performance under the influence of attention.

An initial behavioural rating showed that differences to native sounds are perceivable. The measurement of the pre-attentive and automatic mismatch response confirmed that language background significantly influences the early perception of foreign speech sounds. The direction of the effect, however, was unexpected: Swiss-German speakers displayed higher amplitudes than Rhaeto-Romance speakers on contrasts that are not represented in their native phoneme inventory. There are two possible explanations for this finding. On the hand, a rich second language background in both groups could have evoked memory traces in the Swiss-German speakers as well. Cortical representations of the foreign sound category might have enabled Swiss-German speakers to perform in a comparable if not better way to the Rhaeto-Romance speakers. This would confirm the belief that linguistic experience affects the neural processing window for speech (compare Gandour *et al.*, 2007). On the other hand, overlearning could yield smaller amplitudes to the phonetic contrast in the Rhaeto-Romance group; the higher amplitudes in the Swiss-German subjects would be interpreted as increased neural activity (compare Tervaniemi *et al.*, 2000).

The comparison of two different language groups in their perception of yet another language was relatively unusual. In other studies, either a naive or an advanced group of foreign language learners is compared to native speakers of the language under investigation in their MMN response to phonetic contrasts (e.g. Winkler *et al.*, 1999). Testing the perception abilities of a third language makes the direct comparison of the neural responses difficult as neither of them is a native response.

The naturally spoken Serbian stimuli were stripped of surrounding acoustic cues. This could have played a significant role on the discrimination ability of the Rhaeto-Romance subjects. As Lipski (2006) points out, speakers of languages with an inventory of various fricatives and/or affricates seem to rely more highly on formant transitions than on the frication noise to discriminate these phonemes. Affricate contrasts might need to be placed in their typical context of acoustic cues to be reliably distinguished by native speakers.

Language learning needs to take place in a relevant context – acoustic discrimination ability and knowledge of the related meaning cannot be separated. Phonetic knowledge of the foreign language contrast alone does not enable better discrimination ability if relevant acoustic cues are missing. Thus, it remains unclear whether in our experiment the non-native sounds were assimilated to the native phoneme category or not. Nevertheless, the results support the notion that phonetic features that seem irrelevant to the acquired L1-specific representations are not completely neglected or filtered out. This strongly speaks in favour of the continuous ability to learn foreign language phonemes that are similar/ dissimilar to the L1 phonetic category in adulthood.

6. REFERENCES

Boersma, P. & Weenink, D. (2009), *Praat: doing phonetics by computer*, www.praat.org.

Brunner, R. (1963), Zur Physiologie der Rätoromanischen Affrikaten *tsch* und *tg (ch)*, in *Sprachleben der Schweiz. Sprachwissenschaft, Namenforschung, Volkskunde* (P. Zinsli *et al.*, editors), Bern: Francke, 167-173.

Corbett, G. (1987), Serbo-Croat, in *The world's major languages* (B. Comrie, editor), London: Routledge, 391-409.

Fleischer, J. & Schmid, S. (2006), Zurich German, *Journal of the International Phonetic Association*, 36, 243-253.

Forrest, K., Weismer, G., Milenkovic, P. & R. Dougall (1988), Statistical analysis of word-initial word-final voiceless obstruents: Preliminary data, *Journal of the Acoustical Society of America*, 84, 115-123.

Friederici, A.D., Steinhauer, K. & Pfeifer E. (2002), Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis, *Pnas,* 99, 529-534.

Friederici, A.D. (2005), Neurophysiological markers of early language acquisition: from syllables to sentences, *Trends in Cognitive Sciences,* 9, 481-485.

Gandour, J., Tong, Y., Talavage, T., Wong, D., Dzemidzic, M., Xu, Y. & Lowe, M. (2007), Neural Basis of First and Second Language Processing of Sentence-Level Linguistic Prosody, *Human Brain Mapping,* 28, 94-108.

Gordon, M., Barthmaier, P. & Sands, K. (2002), A cross-linguistic acoustic study of voiceless fricatives*, Journal of the International Phonetic Association,* 32, 141-174.

Grimm, S., Schröger, E., Bendixen, A., Bäss, P., Roye, A. & Deouell, L.Y. (2008), Optimizing the auditory distraction paradigm: Behavioral and event-related potential effects in a lateralized multi-deviant approach, *Clinical Neurophysiology,* 119, 934-947.

Haiman, J. & Benincà, P. (1992), *The Rhaeto-Romance Language,* London: Routledge.

Jaspers, K. (1958), [The physician in the technical age]. *Klin.Wochenschrift,* 36, 1037-1043.

Kuhl, P. (2004), Early Language Acquisition: Cracking the Speech Code, *Nature Revievs,* 5, 831-843.

Ladefoged, P. & Maddieson, I. (1996), *The Sound of the World's Languages*, Oxford: Blackwell.

Lenneberg, E. (1967), *Biological foundations of language*, New York: Wiley.

Lipski, S. C. (2006), *Neural correlates of fricative contrasts across language boundaries,* Institut für Maschinelle Sprachverarbeitung der Universität Stuttgart.

Liver, R. (1999), *Rätoromanisch. Eine Einführung in das Bündnerromanisch,* Tübingen: Narr.

Mele, B. & Schmid, S. (2009), Le occlusive palatali nel dialetto di San Giovanni in Fiore (CS), in *La fonetica sperimentale. Metodo e applicazioni* (L. Romito, V. Galatà & R. Lio, editors), Torriana: EDK, 349-371.

Morén, B. (2006), Consonant-vowel interactions in Serbian: Features, representations and constraint interactions, *Lingua,* 116, 1198-1244.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J. & Alho, K. 1997), Language-specific phoneme representations revealed by electric and magnetic brain responses, *Nature,* 385, 432-434.

Näätänen, R., Pakarinen, S., Rinne, T. & Takegata, R. (2004), The mismatch negativity (MMN): towards the optimal paradigm, *Clinical Neurophysiology,* 115, 140-144.

Nenonen, S., Shestakova, A., Huotilainen, M. & Näätänen, R. (2005), Speech-sound duration processing in a second language is specific to phonetic categories, *Brain and Language*, 92*,* 26-32.

Recasens, D. & Espinosa, A. (2007), An electropalatographic and acoustic study of affricates and fricatives in two Catalan dialects, *Journal of the International Phonetic Association,* 37, 143-172.

Schmid, S. (2010), Les occlusives palatales du Vallader, in *Actes du XXVe Congrès International de Linguistique et Philologie Romanes* (P. Danler, editor), Tübingen: Niemeyer, 185-193.

Tervaniemi, M., Ilvonen, T., Sinkkonen, J., Kujala, A., Alho, K., Huotilainen, M. & Näätänen, R. (2000), Harmonic partials facilitate pitch discrimination in humans: electrophysiological and behavioral evidence, *Neuroscience Letters*, 279, 29-32.

Titova, N. & Näätänen, R. (2001), Preattentive voice discrimination by the human brain as indexed by the mismatch negativity, *Neuroscience Letters,* 308, 63-65.

Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Scepe, V., Ilmoniemi, R. J. & Näätänen, R. (1999), Brain responses reveal the learning of foreign language phonemes, *Psychophysiology,* 36, 638-642.